
ImmuneDB Documentation

Release 0.25.x

Aaron Rosenfeld

Sep 14, 2018

1	Quick Start	3
2	More Information	5
3	Underlying Technologies	7
3.1	Installing with Docker (recommended)	7
3.2	Installing Locally	10
3.3	Running the Data Analysis Pipeline	11
3.4	Command Line Reference	18
3.5	Python API	32
3.6	Directly Querying the Database	40
3.7	Referencing ImmuneDB	41
3.8	Related Publications & Software	42
	Python Module Index	43

ImmuneDB is a system to facilitate efficient storage and analysis of high-throughput B- and T-cell sequence data. It provides V- and J-gene identification, clonal assignment, lineage construction, selection pressure calculation, and thorough exporting functionality.

It also provides an intuitive and useful web interface a demo of which you can see [here](#).

CHAPTER 1

Quick Start

To get started immediately, please see the *Docker installation instructions*.

More Information

ImmuneDB is comprised of two GitHub repositories: Python analysis tools ([arosenfeld/immunedb](#)) and a web interface ([arosenfeld/immunedb-frontend](#))

The system aims to:

- **Reduce ad-hoc scripting:** Data analysis performed on an ad-hoc basis with custom scripts and data formats is error-prone and leads to inconsistencies. ImmuneDB provides a standardized analysis platform, performing many common tasks automatically.
- **Minimize flat-files:** Flat files are currently the standard method of data exchange in the biological sciences. There are a myriad of drawbacks when using these including a lack of [referential integrity](#), unclear [provenance](#), and non-standardized formats.

ImmuneDB attempts to reduce the need for flat files, through the use of an industry-leading database, MySQL. When data must be exchanged as a flat-file, many export options, including FASTA and tab-delineation, are available.

- **Interoperate with existing tools:** ImmuneDB integrates tools from other researchers to provide features such as lineage construction, genotyping, and selection pressure calculations. Further, ImmuneDB can import and export in a variety of common formats, making it compatible with the larger AIRR ecosystem of tools.

ImmuneDB has four primary components:

- **Clonal identification pipeline:** A series of steps which identify probable V- and J-genes from raw FASTA files and subsequently groups sequences from the same subject into clones. At each step, statistics and metadata are stored in the underlying database for later analysis.
- **Analysis framework:** A module that provide detailed analysis of sequences. This includes mutation analysis, gene utilization breakdowns, and feature distributions (e.g. CDR3 length, V-gene utilization).
- **API:** ImmuneDB includes two APIs. First, the REST API allows native and web applications (via AJAX) to interact with the underlying database without writing any database specific code. Further, the REST API is language agnostic, using HTTP to exchange data.

Second, the Python API which implements the data models and querying functionality. This allows developers to write customized queries directly interacting with the database.

- **Web interface:** Distributed separately of ImmuneDB is a web interface which utilizes the REST API to allow for easy browsing of analysis results.

Each of the components can be used independently so long as the underlying database is properly populated.

Underlying Technologies

- **MySQL:** The database (agnostic of implementation like [MariaDB](#)) which enforces rigorous data guarantees, indexes the data for fast retrieval, and enforces atomicity of operations
- **SQLAlchemy:** The interface between Python and the database which abstracts SQL queries from the implementation and enforces the data model within Python.
- **Bottle:** A lightweight web-framework for Python which serves content via the REST API allowing web-applications to issue AJAX requests to the ImmuneDB framework.

3.1 Installing with Docker (recommended)

3.1.1 Dependencies

The only dependency for running ImmuneDB in Docker is to have Docker installed locally.

3.1.2 Pulling the Docker Image

With Docker installed run one of these commands:

```
# Pulls the most recent ImmuneDB version
$ docker pull arosenfeld/immunedb:v0.25.0
```

3.1.3 Running the Container

To start a shell session within the container run:

```
$ docker run -v $HOME/immunedb_share:/share \
  -p 5000:5000 -p 8080:8080 -it arosenfeld/immunedb:v0.25.0
```

This will start a shell with ImmuneDB and accessory scripts pre-installed as well as create a shared directory between the host and Docker container. Files placed in the host's `$HOME/immunedb_share` directory and it will appear in `/share` within the Docker container (and vice versa). Note `$HOME` on macOS is generally `/Users/your_username/` and on Linux it is generally `/home/your_username`.

Additionally, MySQL stores its data in `/share/mysql_data` so databases will persist across multiple container invocations.

The location of important files are:

- `/root/germlines`: IMGT aligned germlines for IGH, TCRA, and TCRB.
- `/apps/clearcut/clearcut`: The `clearcut` executable for generating lineages. This file is in the container's `$PATH`.
- `/usr/local/sbin/serve_immunedb.sh`: A helper script to serve the ImmuneDB web interface. This file is in the container's `$PATH`.
- `/apps/bowtie2/bowtie2`: The local-alignment tool Bowtie2. This file is in the container's `$PATH`.
- `/share/configs`: The recommended directory to store ImmuneDB configurations generated by `immunedb_admin create`.
- `/share/mysql_data`: The location MySQL (specifically MariaDB) will store its data.
- `/example`: A set of example input data to familiarize yourself with ImmuneDB

3.1.4 Running the Example Pipeline

To get started, two input FASTQ files and an associated `metadata.tsv` file are included at `/example`. We recommend running through this example before analyzing your own data to become familiar with the ImmuneDB pipeline.

First, create a database for the example. Note the database root user does not have a password, so we specify a blank one.

```
$ immunedb_admin create example /share/configs/ --admin-pass ""
2018-06-08 17:44:20 [INFO] Creating user "example"
2018-06-08 17:44:20 [INFO] Creating database "example"
2018-06-08 17:44:20 [INFO] Creating config at /share/configs/example.json
2018-06-08 17:44:20 [INFO] Initializing tables
2018-06-08 17:44:21 [INFO] Success!
```

Now that a database has been created, start the V- and J-identification:

```
$ immunedb_identify /share/configs/example.json \
~/germlines/imgt_human_ighv.fasta \
~/germlines/imgt_human_ighj.fasta \
/example
2018-06-08 17:52:28 [INFO] Starting sample Donor1_Colon
# ... output truncated ...
2018-06-08 17:52:33 [INFO] Completed sample Donor1_Spleen in 0.1m - 1458/1470 (99%)
↪identified
```

Then collapse the sequences across the samples:

```
$ immunedb_collapse /share/configs/example.json
2018-06-08 17:58:05 [INFO] Resetting collapse info for subject 1
# ... output truncated ...
2018-06-08 17:58:06 [INFO] Worker 2: Committing collapsed sequences
```

We will then infer clones using the CDR3 similarity method with all default parameters:

```
$ immunedb_clones /share/configs/example.json similarity
2018-06-08 18:00:31 [INFO] Generating task queue for subject 1
# ... output truncated ...
2018-06-08 18:00:34 [INFO] Skipping subclones
```

We then calculate per-sample clone statistics:

```
$ immunedb_clone_stats /share/configs/example.json
2018-06-08 18:01:38 [INFO] Creating task queue to generate stats for 236 clones.
# ... output truncated ...
2018-06-08 18:01:43 [INFO] Worker 2: Clone 236
```

Optionally, we can also generate a lineage for each clone. To reduce the influence of sequencing error, we use `--min-count 2` to include only mutations that occur at least twice:

```
$ immunedb_clone_trees /share/configs/example.json --min-count 2
2018-06-08 15:12:07 [INFO] Creating task queue for clones
# ... output truncated ...
2018-06-08 15:12:08 [INFO] Worker 5: Running clone 236
```

Another optional step is to use `BASELINE` to calculate selection pressure for each clone. Note that this is a relatively slow process, even for this small dataset:

```
$ immunedb_clone_pressure /share/configs/example.json \
  /apps/baseline/Baseline_Main.r
2018-06-08 23:34:32 [INFO] Creating task queue to calculate selection pressure for
↳236 clones.
# ... output truncated ...
2018-06-09 00:35:46 [INFO] Worker 4: Clone 236
```

The last step of the pipeline is to calculate statistics for each sample in the dataset:

```
$ immunedb_sample_stats /share/configs/example.json
2018-06-08 18:04:58 [INFO] Creating task queue to generate stats for sample 1.
# ... output truncated ...
2018-06-08 18:04:59 [INFO] Worker 1: Processing clones for sample 2, include_outliers_
↳False, only_full_reads False
```

At this point the database is fully populated and you can use the web interface and export data.

Using the Web Interface

Now that the database is populated, let's view the data in the web interface using the included helper script. This takes a moment, so wait for the message `webpack: Compiled successfully..`

```
$ serve_immunedb.sh /share/configs/example.json
Running for database /share/configs/example.json
# ... output truncated ...
webpack: Compiled successfully.
```

You should now be able to navigate to `http://localhost:8080` and view the web interface.

Exporting the Data

Finally, lets export the data in [AIRR format](#) and move it to `/share/export` so it is available to the host system:

```
$ mkdir /share/export
$ cd /share/export
$ immunedb_export /share/configs/example.json airr
2018-06-08 18:09:41 [INFO] Exporting subject D1
```

There should now be a `D1.airr.tsv` file in the containers `/share/export` directory and the linked `$HOME/immunedb_share` directory on the host. There is only one file since the AIRR format export breaks the data into one file per subject and this example only has the subject D1.

Conclusion

At this point you've completed the example pipeline. For details on creating your own metadata file and tweaking the pipeline to your needs see [Running the Data Analysis Pipeline](#) and [Command Line Reference](#).

3.2 Installing Locally

This section details how to set ImmuneDB up locally on a machine. This is a more complicated process than using the [Docker method](#) but may be useful if you plan on running ImmuneDB remotely on a server rather than locally.

3.2.1 Dependency Installation

MySQL

ImmuneDB utilizes [MySQL](#) as its underlying data store. We recommend using its drop-in replacement, [MariaDB](#). Please consult their website and your operating systems package manager for installation instructions.

R (optional)

[Baseline](#) can optionally be used to calculate selection pressure on clones. This requires [R](#) to be installed along with the [ade4](#) package. Installation is platform dependent.

The newest version of [Baseline](#) can be downloaded [here](#). The path to the main script will be needed for clone statistics generation as described in [Statistics Generation](#).

For genotyping, [TIgGER](#) must also be installed.

Bowtie2 (optional)

[Bowtie2](#) can be used to locally align sequences which cannot be aligned using the built-in anchor method.

Clearcut (optional)

[Clearcut](#) can be used to generate lineage trees for clones. After downloading and compiling per the instructions, note the path to the `clearcut` executable which will be required for generating trees in [Clone Trees \(Optional\)](#).

3.2.2 ImmuneDB Installation

It is recommended that ImmuneDB be installed within a *venv*, creating an isolated environment from the rest of the system.

To create a virtual environment and activate it run:

```
$ virtualenv immunedb
$ source immunedb/bin/activate
```

Then install ImmuneDB

```
$ pip install immunedb
```

3.3 Running the Data Analysis Pipeline

The primary component of ImmuneDB is its clonal identification pipeline which has the capability to take as input raw sequences, determine likely V and J genes, and finally group similar sequences into clones.

The pipeline is comprised of a number of steps which allows any portion of this process to be replaced by another system. For example, HighV-Quest could be used for V and J assignment portion. Further, the ImmuneDB API allows developers to integrate other tools into each step of the pipeline.

This page explains the basic workflow and assumes MySQL and ImmuneDB are already installed on the system. It does not attempt to detail all the possible options at each stage of the pipeline and users are encouraged to review the *Command Line Reference*.

3.3.1 Quick Start

The details of all ImmuneDB commands are below, however, the following is the basic set of commands to run ImmuneDB. This assumes you have the V-germlines in `imgt_human_v.fasta` and J-germlines in `imgt_human_j.fasta` within the current working directory. It also assumes there is a set of FASTA/FASTQ files in the current working directory.

```
# Navigate to directory with FASTA/FASTQ files
$ cd path/to/my_sequences
# Make a template metadata file
$ immunedb_metadata
# Edit the resulting metadata.tsv file
# Create a database
$ immunedb_admin create example_db ~/configs
# Annotate sequences with V/J genes
$ immunedb_identify ~/configs/example_db.json imgt_human_v.fasta imgt_human_j.fasta .
# Collapse sequences across replicates
$ immunedb_collapse ~/configs/example_db.json
# Assemble clones
$ immunedb_clones ~/configs/example_db.json similarity
# Generate clone statistics
$ immunedb_clone_stats ~/configs/example_db.json
# Generate sample statistics
$ immunedb_sample_stats ~/configs/example_db.json
```

3.3.2 Data Preparation

Before running the ImmuneDB pipeline, the input sequence data must be properly structured. Sequences must be separated into one file per sample. That is, sequences in the same file must be from the sequencing run or, conversely, that sequences in different files could not have originated from the same cell. This is required for ImmuneDB to properly count the number of unique sequences.

For example, a directory of FASTA files may look like this:

```
$ ls
subjectD001_spleen.fasta
subjectD002_blood.fasta
subjectD002_liver.fasta
```

ImmuneDB needs some metadata about each of the FASTA files to process it. Specifically, it **requires** the following information (the maximum number of characters, when applicable, is shown in parenthesis):

- `sample_name` (128): The name of the sample.
- `study_name` (128): The name of study the sample belongs to.
- `subject` (64): A unique identifier for the subject. This must be unique to the entire ImmuneDB instance as they are not contextual to the study. Therefore if two studies use the same identifier for different subjects, they must be given new distinct identifiers.

Any other fields can be added to the metadata and they will be associated with the appropriate samples. Custom fields **must** only contain letters, numbers, and underscores, and must begin with a letter.

This information is specified in a `metadata.tsv` file which must be placed in the same directory as the FASTA files. A template for this file can be generated based on FASTA files with:

A template for the files above could be generated with the following command, while in the same directory:

```
$ immunedb_metadata --use-filenames --include-common
```

Passing `--use-filenames` will automatically populate the `sample_name` field with the name of each file trimmed of its extension. The `--include-common` flag will include some commonly used fields.

The following is an example of such a metadata file with some information filled in:

file_name	study_name	sample_name	subject	date	subset	tissue	disease	lab	experimenter	ig_class	primer1	primer2
subjectD001_SPL.fasta	B-cell Study	subjectD001_SPL	D001	2015-09-13	Naive	Spleen		Some lab	Mr. Experimenter		Leader	J mix
subjectD002_BL.fasta	B-cell Study	subjectD002_BL	D002	2015-09-14	Naive	Blood		Some lab	Mrs. Experimenter		Leader	J mix
subjectD002_Liver.fasta	B-cell Study	subjectD002_Liver	D003	2015-09-15	Ma-ture	Liver		Some lab	Mrs. Experimenter		FW1	J mix

Warning: The values NA, N/A, NULL, and None (case-insensitive) will all be treated as null values in the metadata. Avoid using any other placeholder for missing data, and simply leave the field blank.

After creating the metadata file, the directory should look like:

```
$ ls
metadata.tsv
subjectD001_spleen.fasta
subjectD002_blood.fasta
subjectD003_liver.fasta
```

Germline Files

ImmuneDB requires that V and J germlines be specified in two separate FASTA files. There are a number of restrictions on their format. Most common germlines can be downloaded from [IMGT's Gene-DB](#) directly.

V Germlines

- Genes must be in the format prefixX*Y or prefixX where X is the gene name and Y is the allele. For example, IGHV1-18*01, TRBV5-a*03, and IGHV7-4-1 are all valid. However, IGHV4-34 is not.
- Germlines must be IMGT gapped.
- Germlines starting with gaps are excluded from alignment.
- For anchor identification, ImmuneDB uses the V/J alignment method found in *PMID: 26529062*. This requires V germlines to have one of the following amino-acid anchors with the trailing C being the first residue in the CDR3: D...Y[YHC]C, Y[YHC]C or D...C. The . character represents any amino acid, and [YHC] indicates any one of Y, H, or C. **Local alignment does not place these restrictions on germlines.**

J Germlines

- There must be a fixed number of bases upstream of the CDR3 in all genes.

3.3.3 Main Pipeline

ImmuneDB Instance Creation

It is assumed that the root user's username and password for MySQL is known. To create a new ImmuneDB instance, one can use `immunedb_admin`:

```
$ immunedb_admin create DB_NAME CONFIG_DIR
```

Replacing `DB_NAME` with an appropriate database name and `CONFIG_DIR` with a directory in which the database configuration will be stored will initialize the instance.

Note: By default the root user is used to create the database. You may use a user other than `root` with the `--admin-user` flag, so long as it has permissions to create databases, create users, and grant users permission to manipulate the database in any way.

After running this, a database with the specified name will be created. Further a configuration file with the same name and a `.json` extension will be placed in `CONFIG_DIR`. This configuration file will be the method of referencing the database for the rest of the pipeline steps.

Identifying or Importing Sequences

The first step of the pipeline is to load annotated sequences into the newly created database. This can be done in one of two ways. First, ImmuneDB can take raw FASTA/FASTQ files and use the method proposed by [Zhang, et al.](#) to identify V- and J-genes.

Alternatively, ImmuneDB can take pre-annotated sequences in [Change-O](#) format. The Change-O format was chosen since output from other sources, such as [IMGT High-VQuest](#) and [IgBlast](#) can be easily converted to the Change-O format.

Method 1: Identify Sequences with ImmuneDB

The first step of the pipeline is sequence identification. Primarily this assigns each sequence a V and J gene, but it also calculates statistics such as how well the sequence matches the germline, if there is a probable insertion or deletion, and how far into the CDR3 the V and J likely extend.

```
$ immunedb_identify config.json v_germlines.fasta j_germlines.fasta \  
/path/to/sequence-data-directory
```

Note: J-gene assignment requires three parameters, the number of nucleotides in the J after (upstream) of the CDR3, a conserved anchor size starting at the end of the J, and a minimum anchor length. The J gene is searched for by using these anchors which are 31, 18 and 12 respectively in humans (and are the default values for ImmuneDB). For other species, these values may need to be tweaked. The regions are shown graphically below:

```
                |---- J_MIN_ANCHOR_LEN ----|  
                |----- J_ANCHOR_SIZE -----|  
    ...-- V --|-- CDR3 --|----- J_NTS_UPSTREAM_OF_CDR3 -----|  
j_germline:          ATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG  
seq:                ...ATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG
```

Method 2: Importing from Change-O Format

Importing from the Change-O format is similar to identifying sequences with ImmuneDB.

```
$ immunedb_import config.json v_germlines.fasta j_germlines.fasta \  
/path/to/sequence-data-directory
```

Note that the sequence directory must have a metadata file as described above.

Local Alignment of Indel Sequences (Optional)

Warning: This step is still considered in beta. Some corner cases may not be properly handled, and quality information from FASTQ files will not be included in aligned sequences.

After identification, certain sequences will be marked as being probable indels (or misalignments). To fix these, `immunedb_local_align` can **optionally** be used to properly gap sequences or germlines. It requires `bowtie2` to be installed and in your `PATH`.

```
$ immunedb_local_align config.json v_germlines.fasta j_germlines.fasta
```

Sequence Collapsing

ImmuneDB determines the uniqueness of a sequence both at the sample and subject level. For the latter, `immunedb_collapse` is used to find sequences that are the same except at positions that have an N. Thus, the sequences ATNN and ANCN would be collapsed.

This process has been written in C rather than Python due to its computational complexity. This fact is transparent to the user, however.

To collapse sequences, run:

```
$ immunedb_collapse config.json
```

The optional `--subject-ids` flag can specify that only samples from certain subjects should be collapsed.

Clonal Assignment

After sequences are assigned V and J genes, they can be clustered into clones based on CDR3 Amino Acid similarity with the `immunedb_clones` command. This takes a number of arguments which should be read before use.

There are three ways to create clones: based on CDR3 AA similarity, T-cell exact CDR3 NT identity, and a lineage based method.

Similarity Based

A basic example of similarity-based clonal assignment, not using all possible arguments:

```
$ immunedb_clones config.json similarity
```

This will create clones where all sequences in a clone will have the same V-gene, J-gene, and (by default) 85% CDR3 AA identity.

If you ran local-alignment on sequences, ImmuneDB can also associate clones with insertions or deletions with a probable “parent” clone. The parent clone will have the same V-gene, J-gene, and CDR3 length. Further, the CDR3 amino acid sequences of the subclone will differ by no more than `--min-similarity` (default 85%). This process can be enabled with `--subclones`.

```
$ immunedb_clones config.json --subclones similarity
```

T-cells

If your data is comprised of T-cell sequences, use the T-cell method:

```
$ immunedb_clones config.json tcells
```

This will create clones from the sequences with the same V-gene, J-gene, and identical CDR3 nucleotides.

Lineage Method

Warning: This clone assignment method is still considered in beta.

The lineage based method constructs a lineage for all sequences within subjects that have the same V-gene, J-gene, and CDR3 NT length. It then splits the tree based on common mutations to create clones.

```
$ immunedb_clones config.json lineage
```

Among other arguments, `--mut-cutoff` (default 4) will determine how many mutations must be in common for sequences to be placed in the same clone.

Importing Custom Assignments

If you prefer to import your own clonal assignment, ImmuneDB allows you to export sequences to a file which you can annotate with clone IDs.

```
$ immunedb_clone_import config.json --action export sequences.tsv
```

This will generate a TSV file with all the unique sequences. The last column, `clone_id` will be blank for all rows in the file. To associate sequences together as belonging to a clone, fill in the same value for each of their `clone_id` fields. The value itself can be any string or integer, and only serves as a unique identifier for each clone.

The sequences you assign to a given clone must belong to the same subject and have the same V-gene, J-gene, and number of nucleotides in the CDR3. Further, changing any other values in the TSV file may lead to unpredictable results; they are provided to give adequate information to external clonal assignment programs.

Once the clones have been annotated:

```
$ immunedb_clone_import config.json --action import sequences.tsv
```

Statistics Generation

Two sets of statistics can be calculated in ImmuneDB:

- **Clone Statistics:** For each clone and sample combination, how many unique and total sequences appear as well as the mutations from the germline.
- **Sample Statistics:** Distribution of sequence and clone features on a per-sample basis, including V and J usage, nucleotides matching the germline, copy number, V length, and CDR3 length. It calculates all of these with and without outliers, and including and excluding partial reads.

These are calculated with the `immunedb_clone_stats` and `immunedb_sample_stats` commands and must be run in that order.

```
$ immunedb_sample_stats config.json  
$ immunedb_clone_stats config.json
```

3.3.4 Optional Steps and Features

Genotyping (Optional)

Warning: This step is still considered in beta.

ImmuneDB comes with a helper script to determine the genotype of subjects using [TIgGER](#). This will determine which germline V-genes are present in each subject, and if any contain novel mutations. After this determination, ImmuneDB can operate on the modified genotype FASTA file for further gene identification.

The basic process for this is to identify sequences at the allele level, export sequences in Change-O format, run TIgGer to determine each subject's genotype, delete the originally identified sequences, and then re-run identification with the new V-germlines.

```
$ immunedb_admin create db_name ~/configs
$ immunedb_identify ~/configs/db_name.json v_germlines.fasta j_germlines.fasta \
  /path/to/sequences --genotype
$ immunedb_collapse ~/configs/db_name.json
$ immunedb_export ~/configs/db_name.json changeo --min-subject-copies 2
$ immunedb_genotype ~/configs/db_name.json v_germlines.fasta
$ immunedb_admin delete ~/configs/db_name.json
$ immunedb_admin create db_name ~/configs
# For each subject
$ immunedb_identify ~/configs/db_name.json SUBJECT.v_genotype.fasta j_germlines.fasta \
→ \
  /path/to/SUBJECT_sequence_data
```

Note in the final step (identifying sequences with the inferred genotype) you must specify the sequences only associated with SUBJECT. This step must then be repeated for each subject for which the genotype was inferred.

Selection Pressure (Optional)

Selection pressure of clones can be calculated with [Baseline](#). After installing, and assigning clones with `immunedb_clones`, run:

```
$ immunedb_clone_pressure config.json /path/to/Baseline_Main.r
```

This process is relatively slow and may take some time to complete.

Clone Trees (Optional)

Lineage trees for clones is generated with the `immunedb_clone_trees` command. The only currently supported method is neighbor-joining as provided by [Clearcut](#) which must be in your `$PATH`.

Among others, the `min-count` parameter allows for mutations to be omitted if they have not occurred at least a specified number of times. This can be useful to correct for sequencing error.

```
$ immunedb_clone_trees config.json --min-count 2
```

Web Service (Optional)

ImmuneDB has a RESTful API that allows for language agnostic querying. This is provided by the `immunedb_rest` command. It is specifically designed to provide the required calls for the associated `web-app`.

To run on port 3000 for example:

```
$ immunedb_rest config.json -p 3000
```

3.4 Command Line Reference

This page provides brief descriptions of all the ImmuneDB commands and the default help text for each. For a step-by-step guide to running the pipeline, refer to *Running the Data Analysis Pipeline*.

3.4.1 immunedb_admin

Manages ImmuneDB instances.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_admin", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_admin",
↳line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'
```

immunedb_admin create

Creates a new ImmuneDB instance.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_admin", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
```

(continues on next page)

(continued from previous page)

```

self.require(requires)[0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_admin",
↳line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

immunedb_admin delete

Deletes an ImmuneDB instance.

```

Traceback (most recent call last):
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_admin", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_admin",
↳line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

immunedb_admin backup

Performs a SQL dump from a ImmuneDB instance, backing up the contents.

```

Traceback (most recent call last):
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_admin", line 4, in <module>

```

(continues on next page)

(continued from previous page)

```

__import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_admin",
↳line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

immunedb_admin restore

Restores a SQL dump from `immunedb_admin` backup.

```

Traceback (most recent call last):
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_admin", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_admin",
↳line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_admin')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

3.4.2 immunedb_clone_import

Imports custom clonal assignments from an exported sequence template.

To use, run `immunedb_clone_import ... --action export`, fill in the `clone_id` column for each sequence, then run `immunedb_clone_import ... --action import`. All sequences assigned to a given clone must have the same V-gene, J-gene, and CDR3 length, otherwise an error will be raised.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳ immunedb_clone_import", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clone_import
↳ ')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_clone_
↳ import", line 2, in <module>
    # EASY-INSTALL-SCRIPT: 'ImmuneDB==0.25.0', 'immunedb_clone_import'
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳ 10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳ 12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳ in <module>
ImportError: No module named 'dnautils'
```

3.4.3 immunedb_clone_pressure

Note: For selection pressure calculations, [BASELINE](#) must be installed.

Calculates the selection pressure acting on clonal sequences.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳ immunedb_clone_pressure", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clone_
↳ pressure')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_clone_
↳ pressure", line 4, in <module>
```

(continues on next page)

(continued from previous page)

```

__import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clone_
↳pressure')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

3.4.4 immunedb_clones

Assigns sequences to clones using one of three methods.

```

Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_clones", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clones')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_clones",
↳line 5, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

immunedb_clones similarity

Groups sequences with the same subject, V-gene, J-gene, CDR3 length, and (by default) 85% CDR3 AA similarity into clones.

```

Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_clones", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clones')

```

(continues on next page)

(continued from previous page)

```

File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires) [0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_clones",
↳ line 5, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

immunedb_clones tcells

Groups sequences with the same subject, V-gene, J-gene, and CDR3 NT sequence into clones.

```

Traceback (most recent call last):
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_clones", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clones')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires) [0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_clones",
↳ line 5, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

immunedb_clones lineage

Note: To assign clones via the lineage tree method, [Clearcut](#) must be installed.

Creates lineages out of all sequences with the same subject, V-gene, J-gene, and CDR3 length. Then, the lineage is split along branches where the aggregate number of mutations is at least `--mut-cutoff` (default 4).

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_clones", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clones')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_clones",
↳ line 5, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'
```

3.4.5 `immunedb_clone_stats`

Aggregates statistics about clones for quicker, easier bulk querying.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_clone_stats", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clone_stats')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_clone_
↳stats", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clone_stats')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'
```

3.4.6 immunedb_clone_trees

Note: To create lineage trees, [Clearcut](#) must be installed.

Creates a lineage tree for each clone using Neighbor Joining.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_clone_trees", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clone_trees')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_clone_
↳trees", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_clone_trees')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'
```

3.4.7 immunedb_collapse

Collapses identical sequences across all samples in each subject.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_collapse", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_collapse')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_collapse
↳", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_collapse')
```

(continues on next page)

(continued from previous page)

```

File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

3.4.8 immunedb_export

Exports data from ImmuneDB into various formats

```

Traceback (most recent call last):
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_export", line 4, in <module>
__import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_export')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
self.require(requires)[0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_export",
↳line 4, in <module>
__import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_export')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

immunedb_export changeo

```

Traceback (most recent call last):
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_export", line 4, in <module>
__import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_export')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
self.require(requires)[0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
exec(script_code, namespace, namespace)

```

(continues on next page)

(continued from previous page)

```

File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_export",
↳ line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_export')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

immunedb_export genbank

```

Traceback (most recent call last):
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_export", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_export')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires) [0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_export",
↳ line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_export')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

immunedb_export vjtools

```

Traceback (most recent call last):
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_export", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_export')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires) [0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script

```

(continues on next page)

(continued from previous page)

```

exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_export",
↳ line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_export')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

3.4.9 immunedb_genotype

Note: To genotype subjects, [TIGGER](#) must be installed.

Runs genotyping on a database that was generated with `immunedb_identify ... --genotyping`.

```

Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_genotype", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_genotype')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_genotype
↳", line 7, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

3.4.10 immunedb_identify

Identifies V- and J-genes of sequences in FASTA/FASTQ files using an `anchor method`.

```

Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_identify", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_identify')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_identify
↳", line 2, in <module>
    # EASY-INSTALL-SCRIPT: 'ImmuneDB==0.25.0','immunedb_identify'
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

3.4.11 immunedb_import

Imports sequence alignments and gene calls from Change-O formatted files.

```

Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_import", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_import')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_import",
↳ line 3, in <module>
    __requires__ = 'ImmuneDB==0.25.0'
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

3.4.12 immunedb_local_align

Note: To locally align sequences [Bowtie2](#) must be installed.

Corrects sequences that were flagged as potential indels or unidentifiable by `immunedb_identify`. This can be a slow process for large datasets.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳ immunedb_local_align", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_local_align')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_local_
↳ align", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_local_align')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳ 10, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳ 12, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳ in <module>
ImportError: No module named 'dnautils'
```

3.4.13 immunedb_metadata

Generates metadata for a set of FASTA/FASTQ files to use for `immunedb_identify`.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳ immunedb_metadata", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_metadata')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_metadata
↳ ", line 7, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/identification/__init__.
↳ py", line 3, in <module>
    File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳ python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳ 12, in <module>
```

(continues on next page)

(continued from previous page)

```
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'
```

3.4.14 immunedb_rest

Starts a REST API server for ImmuneDB. This can be used for any purpose, but is designed to provide data for the [ImmuneDB Frontend](#).

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_rest", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_rest')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_rest",
↳line 6, in <module>
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'
```

3.4.15 immunedb_sample_stats

Calculates aggregate statistics for samples for faster querying.

```
Traceback (most recent call last):
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_sample_stats", line 4, in <module>
    __import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_sample_stats
↳')
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
    exec(script_code, namespace)
  File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_sample_
↳stats", line 3, in <module>
```

(continues on next page)

(continued from previous page)

```

__requires__ = 'ImmuneDB==0.25.0'
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

3.4.16 immunedb_sql

Starts an interactive MySQL session for a given ImmuneDB instance. This is simply a wrapper around the `mysql` command that passes information from a configuration file.

```

Traceback (most recent call last):
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/bin/
↳immunedb_sql", line 4, in <module>
__import__('pkg_resources').run_script('ImmuneDB==0.25.0', 'immunedb_sql')
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 654, in run_script
self.require(requires) [0].run_script(script_name, ns)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/pkg_resources/__init__.py", line 1441, in run_script
exec(script_code, namespace, namespace)
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/EGG-INFO/scripts/immunedb_sql",
↳line 7, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/config.py", line
↳10, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/common/models.py", line
↳12, in <module>
File "/home/docs/checkouts/readthedocs.org/user_builds/immunedb/envs/latest/lib/
↳python3.5/site-packages/ImmuneDB-0.25.0-py3.5.egg/immunedb/util/funcs.py", line 4,
↳in <module>
ImportError: No module named 'dnautils'

```

3.5 Python API

Note: This section is currently incomplete. We're working to fill out the details of the Python API as soon as possible.

3.5.1 Configuration

The `immunedb.common.config` module provides methods to initialize a connection to a new or existing database.

Most programs using ImmuneDB will start with code similar to:

```
import immunedb.common.config as config
parser = config.get_base_arg_parser('Some description of the program')
# ... add any additional arguments to the parser ...
args = parser.parse_args()

session = config.init_db(args.db_config)
```

When this script is run, it will require at least one argument which is the path to a database configuration (as generated with `immunedb_admin`). Using that, a `Session` object will be made, connected to the associated database.

One can also directly specify the path to a configuration directly.

```
import immunedb.common.config as config
session = config.init_db('path/to/config')
```

Alternatively a dictionary with the same information can be passed:

```
import immunedb.common.config as config
session = config.init_db({
    'host': '...',
    'database': '...',
    'username': '...',
    'password': '...',
})
```

Returned will be a `Session` object which can be used to interact with the database.

3.5.2 Using the Session

ImmuneDB is built using [SQLAlchemy](#) as a MySQL abstraction layer. Simply put, instead of writing SQL, the database is queried using Python constructs. Full documentation on using the session can be found in [SQLAlchemy's documentation](#).

Once a session is created, the models listed below can be queried.

3.5.3 Example Queries

Below are some example queries that demonstrate how to use the ImmuneDB API.

Clone CDR3s

Get all clones with a given V-gene and print their CDR3 AA sequences.

Input

```
import immunedb.common.config as config
from immunedb.common.models import Clone

session = config.init_db(...)

for clone in session.query(Clone).filter(Clone.v_gene == 'IGHV3-30'):
    print 'clone {} has AAs {}'.format(clone.id, clone.cdr3_aa)
```

Output

```
clone 37884 has AAs CARGYSSSYFDYW
clone 37886 has AAs CARSRSTLSIYGVVPTGDFDSW
clone 37885 has AAs CARNGLNTVSGVVISPQYWLDPW
clone 37887 has AAs CARDLFRGVDFYYGMDVW
```

Clone Frequency

Determine how many sequences appear in each sample belonging to clone 1234.

Note the CloneStats model has one entry for each clone/sample combination plus one where the sample_id field is null which represents the overall clone.

Input

```
import immunedb.common.config as config
from immunedb.common.models import CloneStats

session = config.init_db(...)
for stat in session.query(CloneStats).filter(
    CloneStats.clone_id == 1234).order_by(CloneStats.sample_id):
    print 'clone {} has {} unique sequences and {} copies {}'.format(
        stat.clone_id,
        stat.unique_cnt,
        stat.total_cnt,
        ('in sample ' + stat.sample.name) if stat.sample else 'overall')
```

Output

```
clone 1234 has 53 unique sequences and 1331 copies overall
clone 1234 has 27 unique sequences and 379 copies in sample sample1
clone 1234 has 27 unique sequences and 339 copies in sample sample3
clone 1234 has 24 unique sequences and 311 copies in sample sample4
clone 1234 has 28 unique sequences and 302 copies in sample sample10
```

V-gene Usage

This is a more complex query which gathers the V-gene usage of all sequences which are (a) in subject with ID 1, (b) associated with a clone, and (c) are unique to the subject, printing them from least to most frequent.

Input

```
import immunedb.common.config as config
from immunedb.common.models import Sequence, SequenceCollapse

session = config.init_db(...)

subject_unique_seqs = session.query(
    func.count(Sequence.seq_id).label('count'),
    Sequence.v_gene
).join(
    SequenceCollapse
).filter(
    Sequence.subject_id == 1,
    ~Sequence.clone_id.is_(None),
    SequenceCollapse.copy_number_in_subject > 0
```

(continues on next page)

(continued from previous page)

```

).group_by(
    Sequence.v_gene
).order_by(
    'count'
)

for seq in subject_unique_seqs:
    print seq.v_gene, seq.count

```

Output

```

# ... output truncated ...
IGHV4-34 1128
IGHV1-2 1160
IGHV3-48 1169
IGHV4-39 1310
IGHV3-7 1345
IGHV3-30|3-30-5|3-33 1607
IGHV3-23|3-23D 1626
IGHV3-21 1878

```

3.5.4 Data Models

class `immunedb.common.models.Clone` (**kwargs)
 A group of sequences likely originating from the same germline

Parameters

- **id** (*int*) – An auto-assigned unique identifier for the clone
- **functional** (*bool*) – If the clone is functional
- **v_gene** (*str*) – The V-gene assigned to the sequence
- **j_gene** (*str*) – The J-gene assigned to the sequence
- **cdr3_nt** (*str*) – The consensus nucleotides for the clone
- **cdr3_num_nts** (*int*) – The number of nucleotides in the group’s CDR3
- **cdr3_aa** (*str*) – The amino-acid sequence of the group’s CDR3
- **subject_id** (*int*) – The ID of the subject to which the clone belongs
- **subject** (*Relationship*) – Reference to the associated *Subject* instance
- **germline** (*str*) – The germline sequence for this clone
- **tree** (*str*) – The textual representation of the clone’s lineage tree
- **parent_id** (*int*) – The (possibly null) ID of the clone’s parent

consensus_germline

Returns the consensus germline for the clone

deletions

Returns the list of deletion position/length pairs

insertions

Returns the list of insertion position/length pairs

regions

Returns the IMGT region boundaries for the clone

class `immunedb.common.models.CloneStats` (**kwargs)

Stores statistics for a given clone and sample. If sample is null the statistics are for the specified clone in all samples.

Parameters

- **clone_id** (*int*) – The clone ID
- **clone** (*Relationship*) – Reference to the associated *Clone* instance
- **functional** (*bool*) – If the associated clone is functional. This is a denormalized field.
- **sample_id** (*int*) – The sample ID
- **sample** (*Relationship*) – Reference to the associated *Sample* instance
- **unique_cnt** (*int*) – The number of unique sequences in the clone in the sample
- **total_cnt** (*int*) – The number of total sequences in the clone in the sample
- **mutations** (*str*) – A JSON stanza of mutation count information

class `immunedb.common.models.ModificationLog` (**kwargs)

A log message for a database modification

Parameters

- **id** (*int*) – The ID of the log message
- **datetime** (*datetime*) – The date and time of the message
- **action_type** (*str*) – A short string representing the action
- **info** (*str*) – A JSON stanza with log message information

class `immunedb.common.models.NoResult` (**kwargs)

A sequence which could not be match with a V or J.

Parameters

- **pk** (*int*) – A primary key for this no result
- **seq_id** (*str*) – A unique identifier for the sequence as output by the sequencer
- **sample_id** (*int*) – The ID of the sample from which this sequence came
- **sample** (*Relationship*) – Reference to the associated *Sample* instance
- **sequence** (*str*) – The sequence of the non-identifiable input
- **sequence** – The quality of the non-identifiable input

class `immunedb.common.models.Sample` (**kwargs)

A sample of sequences.

Parameters

- **id** (*int*) – An auto-assigned unique identifier for the sample
- **name** (*str*) – A unique name for the sample as defined by the experimenter
- **study_id** (*int*) – The ID of the study under which the subject was sampled
- **study** (*Relationship*) – Reference to the associated *Study* instance
- **subject_id** (*int*) – The ID of the subject from which the sample was taken

- **subject** (*Relationship*) – Reference to the associated *Subject* instance
- **v_ties_mutations** (*float*) – Average mutation rate of sequences in the sample
- **v_ties_len** (*float*) – Average length of sequences in the sample

class `immunedb.common.models.SampleMetadata` (**kwargs)

class `immunedb.common.models.SampleStats` (**kwargs)

Aggregate statistics for a sample. This exists to reduce the time queries take for a sample.

Parameters

- **sample_id** (*int*) – The ID of the sample for which the statistics were generated
- **sample** (*Relationship*) – Reference to the associated *Sample* instance
- **filter_type** (*str*) – The type of filter for the statistics (e.g. functional)
- **outliers** (*bool*) – If outliers were included in the statistics
- **full_reads** (*bool*) – If only full reads were included in the statistics
- **v_identity_dist** (*str*) – Distribution of V gene identity
- **v_match_dist** (*str*) – Distribution of V gene match count
- **v_length_dist** (*str*) – Distribution of V gene total length
- **j_match_dist** (*str*) – Distribution of J gene match count
- **j_length_dist** (*str*) – Distribution of J gene total length
- **v_gene_dist** (*str*) – Distribution of V-gene assignments
- **j_gene_dist** (*str*) – Distribution of J-gene assignments
- **copy_number_dist** (*str*) – Distribution of copy number
- **cdr3_length_dist** (*str*) – Distribution of CDR3 lengths
- **sequence_cnt** (*int*) – The total number of sequences
- **in_frame_cnt** (*int*) – The number of in-frame sequences
- **stop_cnt** (*int*) – The number of sequences containing a stop codon
- **functional_cnt** (*int*) – The number of functional sequences
- **no_result_cnt** (*int*) – The number of invalid sequences

class `immunedb.common.models.SelectionPressure` (**kwargs)

class `immunedb.common.models.Sequence` (**kwargs)

Represents a single unique sequence.

Parameters

- **ai** (*int*) – An auto-incremented value for the sequence
- **subject_id** (*int*) – The ID of the subject for this subject
- **seq_id** (*str*) – A unique identifier for the sequence as output by the sequencer
- **sample_id** (*int*) – The ID of the sample from which this sequence came
- **sample** (*Relationship*) – Reference to the associated *Sample* instance
- **partial** (*bool*) – If the sequence is a partial read

- **probable_indel_or_misalign** (*bool*) – If the sequence likely has an indel or is a bad alignment
- **v_gene** (*str*) – The V-gene assigned to the sequence
- **j_gene** (*str*) – The J-gene assigned to the sequence
- **num_gaps** (*int*) – Number of inserted gaps within the V read
- **seq_start** (*int*) – The offset from the germline where the sequence starts
- **v_match** (*int*) – The number of V-gene nucleotides matching the germline
- **v_length** (*int*) – The length of the V-gene segment prior to a streak of mismatches in the CDR3
- **j_match** (*int*) – The number of J-gene nucleotides matching the germline
- **j_length** (*int*) – The length of the J-gene segment after a streak of mismatches in the CDR3
- **removed_prefix** (*str*) – The sequence (if any) which was removed from the beginning of the sequence during alignment. Possibly used during indel correction
- **removed_prefix_qual** (*str*) – The quality (if any) which was removed from the beginning of the sequence during alignment. Possibly used during indel correction
- **pre_cdr3_length** (*int*) – The length of the V-gene prior to the CDR3
- **pre_cdr3_match** (*int*) – The number of V-gene nucleotides matching the germline prior to the CDR3
- **post_cdr3_length** (*int*) – The length of the J-gene after to the CDR3
- **post_cdr3_match** (*int*) – The number of J-gene nucleotides matching the germline after to the CDR3
- **in_frame** (*bool*) – If the sequence's CDR3 has a length divisible by 3
- **functional** (*bool*) – If the sequence is functional
- **stop** (*bool*) – If the sequence contains a stop codon
- **copy_number** (*int*) – Number of reads in the sample which collapsed to this sequence
- **cdr3_num_nts** (*int*) – The number of nucleotides in the CDR3
- **cdr3_nt** (*str*) – The nucleotides comprising the CDR3
- **cdr3_aa** (*str*) – The amino-acids comprising the CDR3
- **sequence** (*str*) – The (possibly-padded) sequence
- **quality** (*str*) – Optional Phred quality score (in Sanger format) for each base in sequence
- **germline** (*str*) – The germline sequence for this sequence
- **clone_id** (*int*) – The clone ID to which this sequence belongs
- **clone** (*Relationship*) – Reference to the associated *Clone* instance
- **mutations_from_clone** (*str*) – A JSON stanza with mutation information

clone_sequence

Gets the sequence within the context of the associated clone by adding insertions from other sequences to this one.

deletions

Returns the list of deletion position/length pairs

get_v_extent (*in_clone*)

Returns the estimated V length, including the portion in the CDR3

insertions

Returns the list of insertions position/length pairs

original_quality

Returns the original quality given with the J end trimmed to the germline

original_sequence

Returns the original sequence given with the J end trimmed to the germline

regions

Returns the IMGT region boundaries for the sequence

class `immunedb.common.models.SequenceCollapse` (**kwargs)

A one to many table that links sequence from different samples that collapse to one another. This is used instead of a field in `Sequence` for performance reasons.

Parameters

- **sample_id** (*int*) – The ID of the sample with the sequence being collapsed
- **seq_ai** (*int*) – The auto-increment value of the sequence being collapsed
- **clone** (*Relationship*) – Reference to the associated `Sequence` instance being collapsed
- **collapse_to_subject_sample_id** (*int*) – The ID of the sample in which the collapse to sequence belongs
- **collapse_to_subject_seq_ai** (*int*) – The auto-increment value of the sequence collapsing to
- **collapse_to_subject_seq_id** (*int*) – The sequence ID of the sequence collapsing to. This is a denormalized field.
- **instances_in_subject** (*int*) – The number of instance of the sequence in the subject
- **copy_number_in_subject** (*int*) – The aggregate copy number of the sequence in the subject

collapse_to_seq

Returns the sequence being collapse to

class `immunedb.common.models.Study` (**kwargs)

A study which aggregates related samples.

Parameters

- **id** (*int*) – An auto-assigned unique identifier for the study
- **name** (*str*) – A unique name for the study
- **info** (*str*) – Optional information about the study

class `immunedb.common.models.Subject` (**kwargs)

A subject which was sampled for a study.

Parameters

- **id** (*int*) – An auto-assigned unique identifier for the subject

- **identifier** (*str*) – An identifier for the subject as defined by the experimenter
- **study_id** (*int*) – The ID of the study under which the subject was sampled
- **study** (*Relationship*) – Reference to the associated *Study* instance

`immunedb.common.models.check_string_length(cls, key, inst)`

Checks if a string can properly fit into a given field. If it is too long, a `ValueError` is raised. This prevents MySQL from truncating fields that are too long.

3.6 Directly Querying the Database

ImmuneDB is backed by a MySQL database that can be queried directly to gather information, bypassing the Python API.

3.6.1 Accessing the Database

There are many ways to access the database directly. The two introduced here are directly through MySQL or using `immunedb_sql` which simply wraps a call to MySQL.

With the `immunedb_sql` wrapper (recommended)

```
$ immunedb_sql PATH_TO_CONFIG
```

This is entirely equivalent to using `mysql` and will drop to the MySQL interpreter. You can also pass a query directly from the command line. For example:

```
$ immunedb_sql PATH_TO_CONFIG --query 'select * from samples'
```

Directly with MySQL

From the command line, you may access an ImmuneDB database `DATABASE` from user `USERNAME` with:

```
$ mysql -u USERNAME -p DATABASE
```

This will prompt for a password and then to the database. This method of access is useful for quickly querying the database. To save results of a query `QUERY` run the command:

```
$ mysql -u USERNAME -p DATABASE -e "QUERY" > output
```

Since ImmuneDB stores usernames and passwords in config files `immunedb_sql` is provided as a small wrapper around the `mysql` command. It can be invoked with:

3.6.2 Querying

The data is split into different referential tables based on the data models defined in *Models*.

Most tables contain data that will generally be aggregated by a query directly. For example, to determine how many sequences are in sample 10, one could simply count the rows:

```
> SELECT COUNT(*) FROM sequences WHERE sample_id=10;
```

This query will work, but may be slow with large datasets. Two tables can assist in some computationally-expensive queries by providing pre-aggregated information. The CloneStats and SampleStats tables are pre-populated with the `immunedb_clone_stats` and `immunedb_sample_stats` *pipeline* commands.

This same result could be achieved more quickly with:

```
> SELECT sequence_cnt FROM sample_stats WHERE filter_type='all' AND
outlier=1 AND full_reads=0 AND sample_id=10;
```

This is more verbose but is quicker and also makes other more complex tasks simpler. For example, let's find how many unique sequences are in sample 10 which are full reads and not outliers:

```
> SELECT sequence_cnt FROM sample_stats WHERE filter_type='unique' AND
outlier=0 AND full_reads=1 AND sample_id=10;
```

The CloneStats table also has useful information that has been pre-aggregated. For example, how many unique sequences in sample 10 are in clone 5:

```
> SELECT unique_cnt FROM clone_stats WHERE sample_id=10 AND clone_id=5;
```

Or how many total unique sequences are in clone 5 (NULL is a placeholder in the `sample_id` column meaning "All Samples"):

```
> SELECT unique_cnt FROM clone_stats WHERE sample_id=NULL AND clone_id=5;
```

3.6.3 Other Example Queries

How many clones have a CDR3 starting with 'CARD'?

```
> SELECT COUNT(*) FROM clones WHERE clones.cdr3_aa LIKE 'CARD%';
```

Get a list of non-identifiable sequences in FASTA format.

```
> SELECT CONCAT('>', seq_id, '\n', sequence) FROM noresults;
```

How many indels and total sequences do I have?"

```
> SELECT SUM(IF(probable_indel_or_misalign=1, 1, 0)) AS indels, COUNT(*) AS
total FROM sequences;
```

3.7 Referencing ImmuneDB

If you use ImmuneDB in your work please cite it as:

Rosenfeld, A.M., Meng, W., Luning Prak, E.T., Hershberg, U. (2017) **ImmuneDB: A system for the analysis and exploration of high-throughput adaptive immune receptor sequencing data**. *Bioinformatics*, 33, 292.

3.8 Related Publications & Software

Clumpiness: Schwartz, G.W., Shokoufandeh, A., Ontanon, S., Hershberg, U. (2016) **Using a novel clumpiness measure to unite data with metadata: finding common sequence patterns in immune receptor germline V genes.** Pattern Recognition Letters 74: 24-29

ConservedIdentification: Zhang, B., Meng, W., Luning Prak, E.T. and Hershberg U. (2015) **Discrimination of germline V genes at different sequencing lengths and mutational burdens: A new tool for identifying and evaluating the reliability of V gene assignment.** Journal of Immunological Methods 427: 105-116

Diversity: Schwartz, G.W., Hershberg, U. (2013) **Conserved variation: identifying patterns of stability and variability in BCR and TCR V genes with different diversity and richness metrics.** Phys Biol 10

i

`immunedb.common.models`, 35

C

check_string_length() (in module immunedb.common.models), 40

Clone (class in immunedb.common.models), 35

clone_sequence (immunedb.common.models.Sequence attribute), 38

CloneStats (class in immunedb.common.models), 36

collapse_to_seq (immunedb.common.models.SequenceCollapse attribute), 39

consensus_germline (immunedb.common.models.Clone attribute), 35

D

deletions (immunedb.common.models.Clone attribute), 35

deletions (immunedb.common.models.Sequence attribute), 38

G

get_v_extent() (immunedb.common.models.Sequence method), 39

I

immunedb.common.models (module), 35

insertions (immunedb.common.models.Clone attribute), 35

insertions (immunedb.common.models.Sequence attribute), 39

M

ModificationLog (class in immunedb.common.models), 36

N

NoResult (class in immunedb.common.models), 36

O

original_quality (immunedb.common.models.Sequence attribute), 39

original_sequence (immunedb.common.models.Sequence attribute), 39

R

regions (immunedb.common.models.Clone attribute), 35

regions (immunedb.common.models.Sequence attribute), 39

S

Sample (class in immunedb.common.models), 36

SampleMetadata (class in immunedb.common.models), 37

SampleStats (class in immunedb.common.models), 37

SelectionPressure (class in immunedb.common.models), 37

Sequence (class in immunedb.common.models), 37

SequenceCollapse (class in immunedb.common.models), 39

Study (class in immunedb.common.models), 39

Subject (class in immunedb.common.models), 39